

CLAIMS

What is claimed is:

SUB C47

10014744-102601

1. A method for classification of data, comprising:
  - providing first data from a physical process, the first data including a class label associated with attributes of the first data;
  - deriving a dummy cluster from centroid coordinates of the first data associated with the class label;
  - determining distance measures between the first data and a plurality of clusters which include the dummy cluster;
  - creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster;
  - identifying a closest match between second data and the plurality of clusters; and
  - classifying the second data based on a class label of the closest match from the plurality of clusters.
2. The method of claim 1, wherein deriving a dummy cluster includes:
  - creating first and second dummy clusters associated with first and second class labels, respectively;
  - initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and
  - initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

3. The method of claim 2, wherein deriving a dummy cluster further includes:

updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

4. The method of claim 1, wherein determining distance measures between the first data and a plurality of clusters includes using one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance for the distance measure.

5. The method of claim 1, wherein determining distance measures between the first data and a plurality of clusters includes:

calculating sample variance of the first data;

calculating sample covariance of the first data;

calculating sample mean of the first data; and

calculating correlation coefficient from the sample variance, sample covariance, and sample mean of the first data.

6. The method of claim 1, further including creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

10014744-102601

7. The method of claim 1, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

8. A method of classifying first data from a physical process, comprising:

providing first data which includes a class label associated with attributes of the first data;

deriving a dummy cluster from centroid coordinates of the first data associated with the class label;

determining distance measures between the first data and a plurality of clusters which include the dummy cluster; and

creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster.

9. The method of claim 8 further including:

identifying a closest match between second data and the plurality of clusters; and

classifying the second data based on a class label of the closest match from the plurality of clusters.

10. The method of claim 9, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

10014744-102601

11. The method of claim 8, wherein deriving a dummy cluster includes:

creating first and second dummy clusters associated with first and second class labels, respectively;

initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

12. The method of claim 11, wherein deriving a dummy cluster further includes:

updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

13. The method of claim 8, further including creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

10014744-102601

14. A method of classifying first data from a physical process, comprising:

providing first data which includes a class label associated with attributes of the first data;  
deriving a dummy cluster from centroid coordinates of the first data associated with the class label;  
determining distance measures between the first data and a plurality of clusters which include the dummy cluster; and  
creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

15. The method of claim 14 further including:

identifying a closest match between second data and the plurality of clusters; and  
classifying the second data based on a class label of the closest match from the plurality of clusters.

16. The method of claim 15, wherein identifying a closest match between second data and the plurality of clusters includes calculating a distance measure from one of the group of a weighted Canberra distance, a weighted Euclidean distance, and a weighted Chi-squared distance and using the closest distance measure as the closest match.

10014744-102501

17. The method of claim 14, wherein deriving a dummy cluster includes:

creating first and second dummy clusters associated with first and second class labels, respectively;

initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

18. The method of claim 17, wherein deriving a dummy cluster further includes:

updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

19. The method of claim 14, further including creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster.

10014744-102601

20. A digital storage medium encoded with a computer program which classifies data, the computer program comprising:

first instructions for providing first data from a physical process, the first data including a class label associated with attributes of the first data;

second instructions for deriving a dummy cluster from centroid coordinates of the first data associated with the class label;

third instructions for determining distance measures between the first data and a plurality of clusters which include the dummy cluster; and

fourth instructions for creating a real cluster in the plurality of clusters if the first data is closest to the dummy cluster.

21. The digital storage medium of claim 20 further including:

fifth instructions for identifying a closest match between second data and the plurality of clusters; and

sixth instructions for classifying the second data based on a class label of the closest match from the plurality of clusters.

22. The digital storage medium of claim 20, wherein the second instructions further include:

fifth instructions for creating first and second dummy clusters associated with first and second class labels, respectively;

sixth instructions for initializing the first dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the first class label; and

seventh instructions for initializing the second dummy cluster with the centroid coordinates of the first data if the class label of the first data matches the second class label.

100-14744-102501

23. The digital storage medium of claim 22, wherein the second instructions further include:

    eighth instructions for updating the centroid coordinates of the first dummy cluster for each first data having a class label that matches the first class label; and

    ninth instructions for updating the centroid coordinates of the second dummy cluster for each first data having a class label that matches the second class label.

24. The digital storage medium of claim 20, wherein the second instructions further include creating a real cluster in the plurality of clusters if the first data is closest to a cluster having a class label different than the class label associated with the first data.

10014744-102601